

1/10

Systematic Gene Search in the Incyte LifeSeq Database

Normal tissue
~50,000 individual ESTs

Tumor tissue
~50,000 individual ESTs

Priority list
High

Prostate
Breast
Ovary
Bladder
Uterus

Iterative assembling
with
increasing mismatch

Low

~8,000 contigs
+
~25,000 individual
sequences

~8,000 contigs
+
~25,000 individual
sequences

Comparison of databases

normal tissue-
specific
(expected: 100-500)

nonspecifically
expressed genes

tumor tissue-
specific
(expected: 100-500)

Genes of Interest

Figure 1

CORRECTED PAGE (RULE 91)
ISA/EP

1/10

Systematische Gen-Suche in der Incyte LifeSeq Datenbank

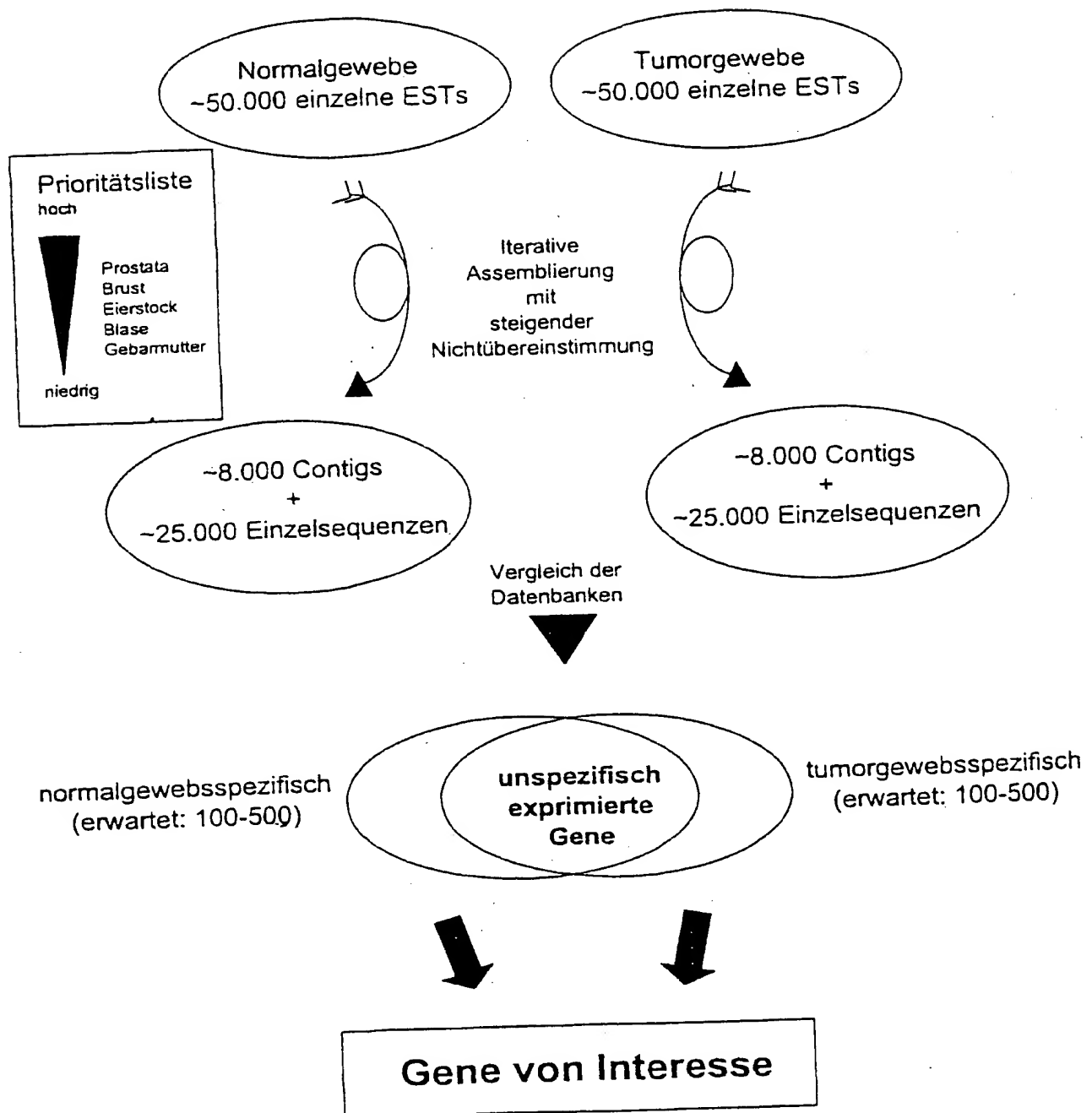


Fig. 1

2/10

Principle of EST Assembly

~50,000 ESTs per tissue

Assembly at 0% mismatch
with GAP4 (Staden)

Contigs

Individual Sequences

Contigs increasing in
number and lengthIterative assembly with
increasing mismatch
(1%, 2%, 4%)

5000-6000 contigs

~25,000 other individual
sequences~30,000 consensus-
sequences per tissue

Figure 2a

CORRECTED PAGE (RULE 91)
ISA/EP

2/10

Prinzip der EST-Assemblierung

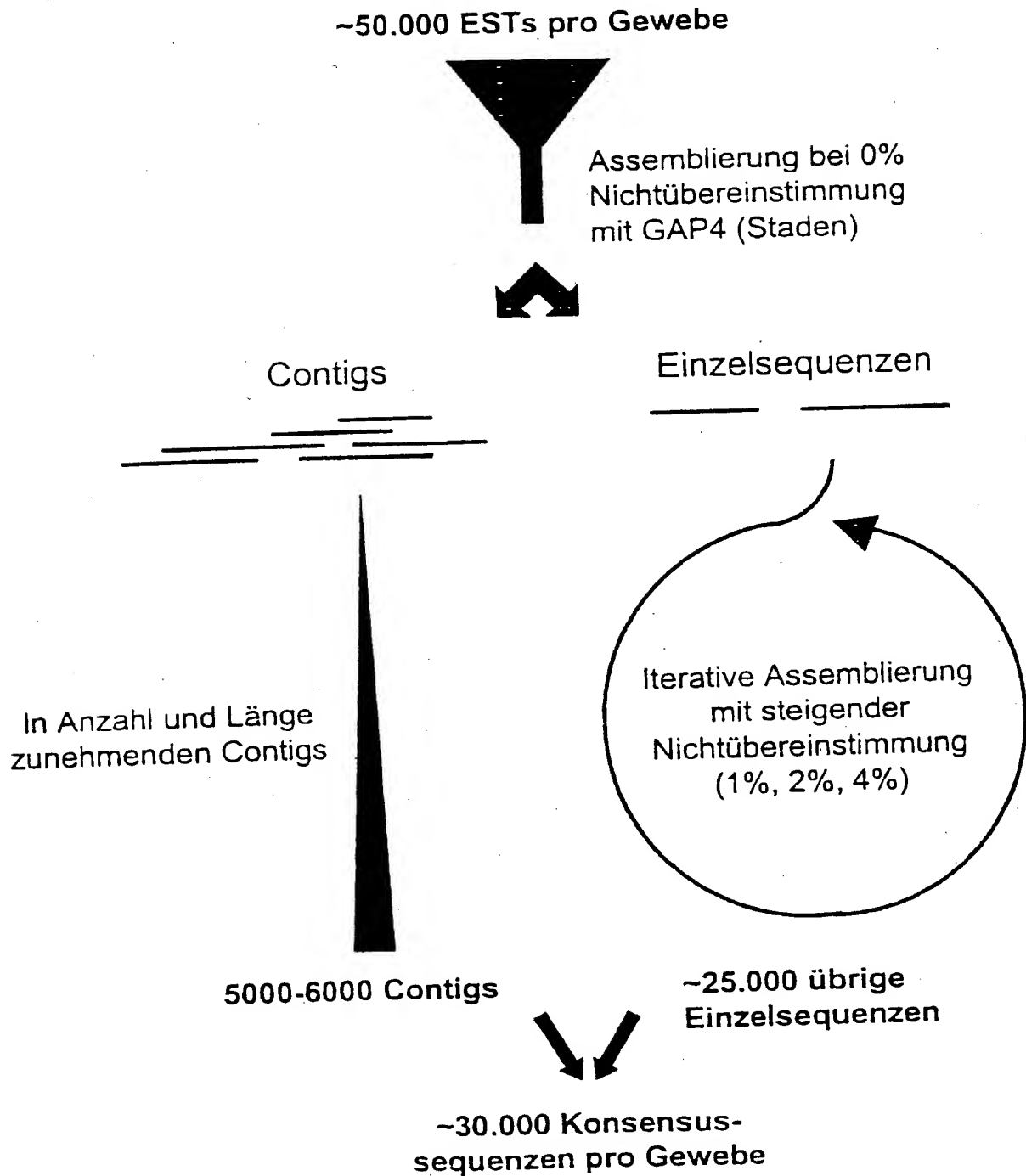


Fig. 2a

3/10

~50,000 ESTs of a tissue (e.g.: uterus tumor)

GAP4 Assembly 1st Round:
minimum initial match: 20
maximum number of inserted blanks per sequence: 8
maximum percent mismatch: 0

GAP4 Database 1
Contigs 1
Individual sequences 1

unassembled
ESTs

GAP4 Assembly 2nd Round:
minimum initial match: 20
maximum number of inserted blanks per sequence: 8
maximum percent mismatch: 1

GAP4 Database 2
Contigs 2
Individual sequences 2

unassembled
ESTs

GAP4 Assembly 3rd Round:
minimum initial match: 20
maximum number of inserted blanks per sequence: 8
maximum percent mismatch: 2

GAP4 Database 3:
Contigs 3
Individual sequences 3

unassembled
ESTs

Figure 2b1

CORRECTED PAGE (RULE 91)
ISA/EP

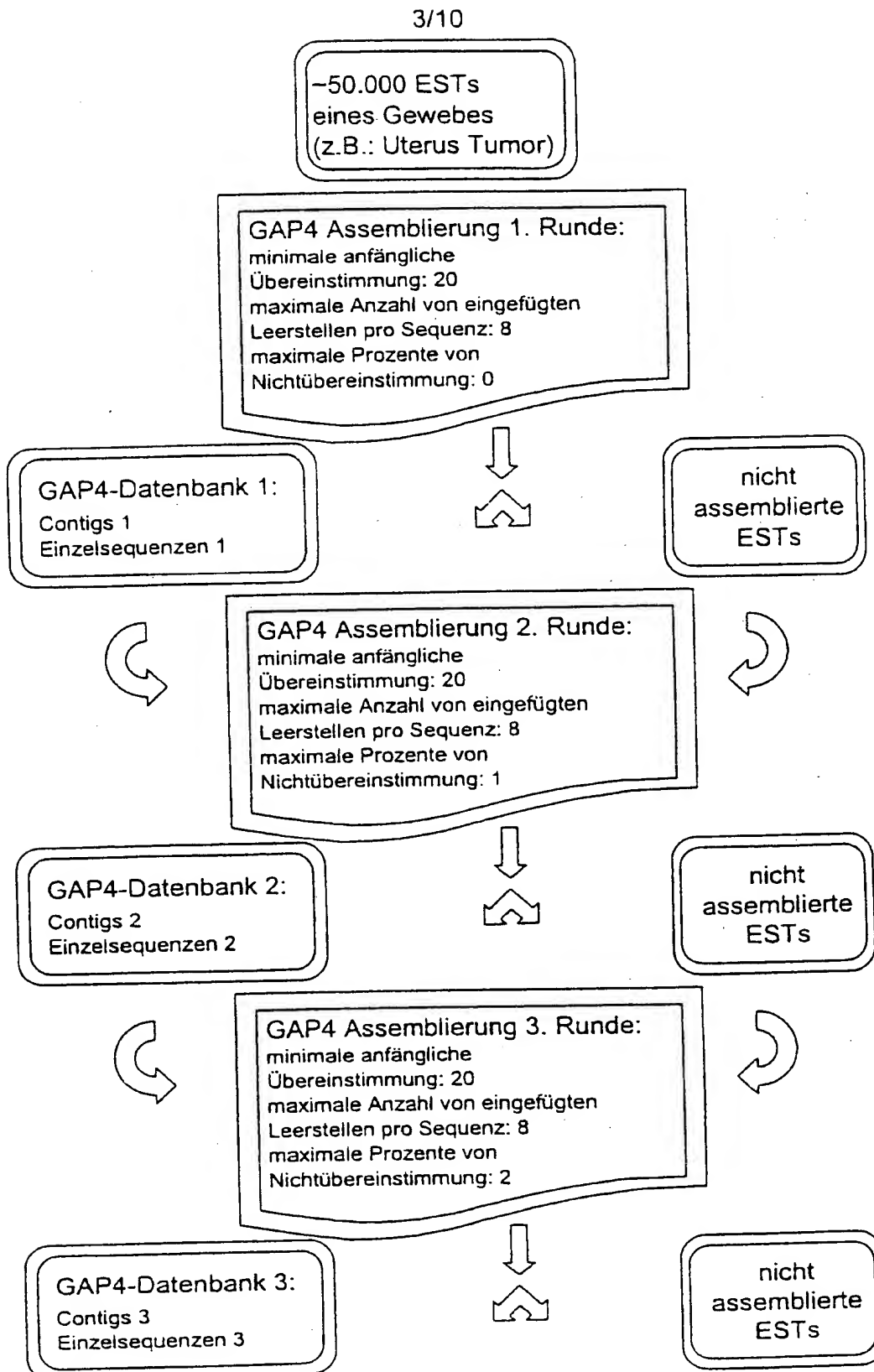


Fig. 2b1

unassembled
ESTs

GAP4 Assembly 4th Round:
minimum initial match: 20
maximum number of inserted blanks
per sequence: 8
maximum percent mismatch: 2

unassembled
ESTs

```
GAP4 Assembly 5th Round:
minimum initial match: 20
maximum number of inserted blanks
per sequence: 8
maximum percent mismatch: 4
```

unassembled
ESTs 5

Individual Sequences 5

Figure 2b2

CORRECTED PAGE (RULE 91)
ISA/EP

4/10

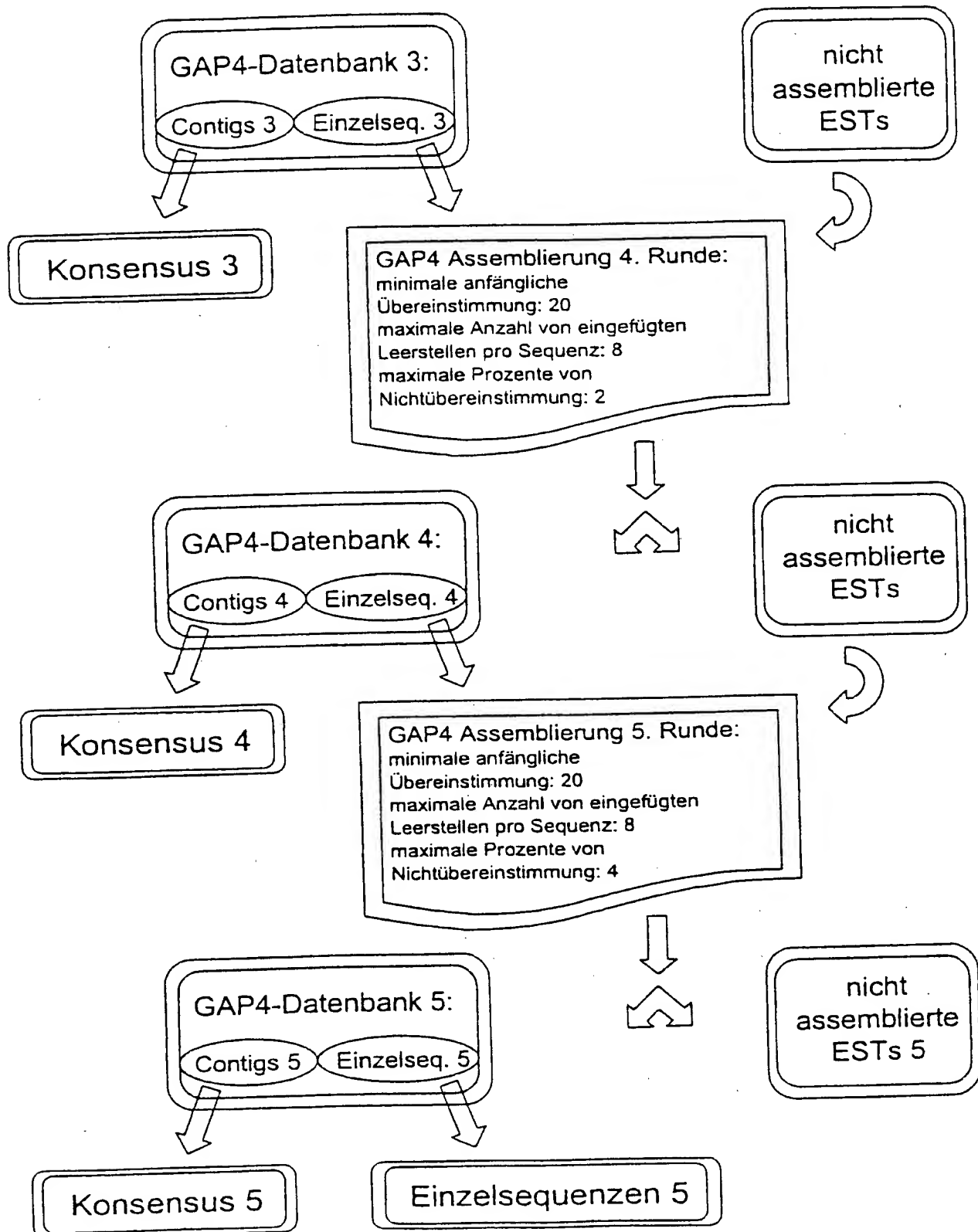


Fig. 2b2

5/10

Consensus 3.

Individual Sequences 5

Consensus 4

unassembled
ESTs 5

Consensus 5

GAP4 Assembly 6th Round:
minimum initial match: 20
maximum number of inserted blanks per sequence: 8
maximum percent mismatch: 4

Assembled database
of a specific tissue
(e.g.: uterus tumor)

Figure 2b3

CORRECTED PAGE (RULE 91)
ISA/EP

5/10

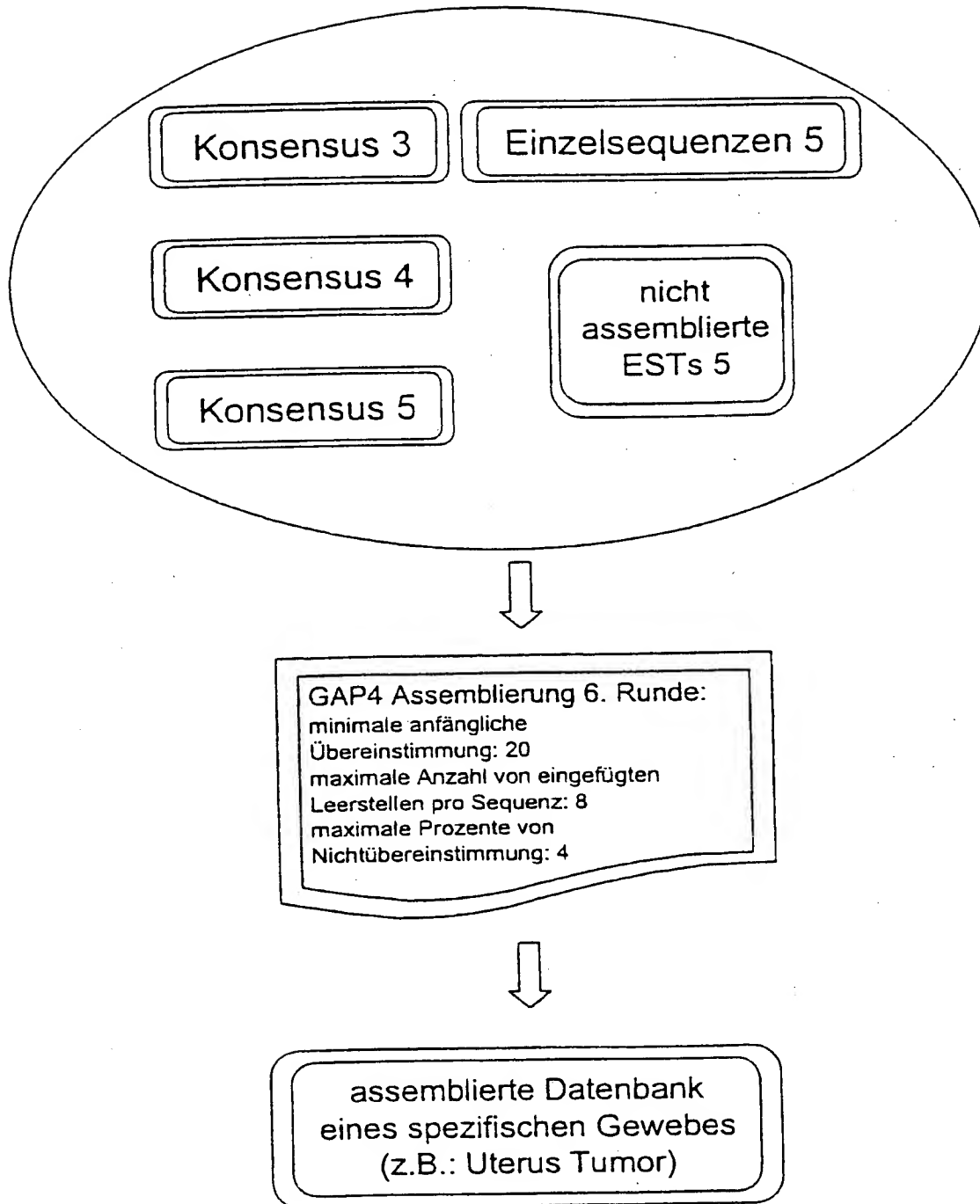


Fig. 2b3

6/10

Assembled database
of a specific tissue
(e.g.: uterus tumor)

Consensus 6

Read-in as individual sequences

Database
of a specific tissue
(e.g.: uterus tumor)

Database of a second
specific tissue
(e.g.: normal uterus)

GAP4 Assembly
minimum initial match: 20
maximum number of inserted blanks
per sequence: 8
maximum percent mismatch: 4

Tumor tissue-
specific ESTs

Non-tissue-
specific ESTs

Normal tissue-
specific ESTs

Fig. 2b4

CORRECTED PAGE (RULE 91)
ISA/EP

6/10

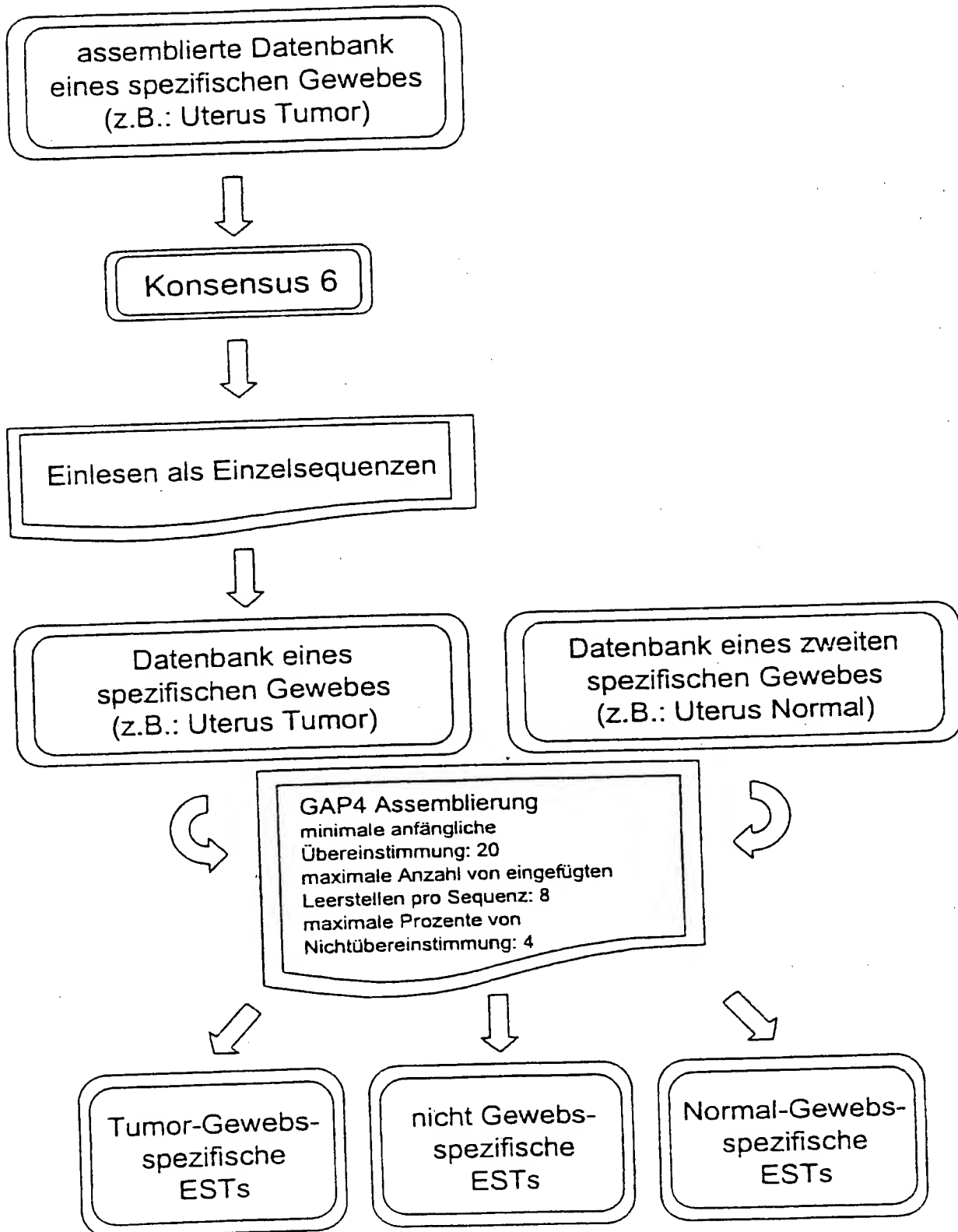


Fig. 2b4

7/10

In silico subtraction of gene expression in various tissues

~30,000 consensus sequences
normal tissue

~30,000 consensus sequences
tumor tissue

Assembly at 4% mismatch

Normal tissue
Specific genes

Cancer tissue
Specific genes

Genes expressed in both tissues

Figure 3

CORRECTED PAGE (RULE 91)
ISA/EP

7/10

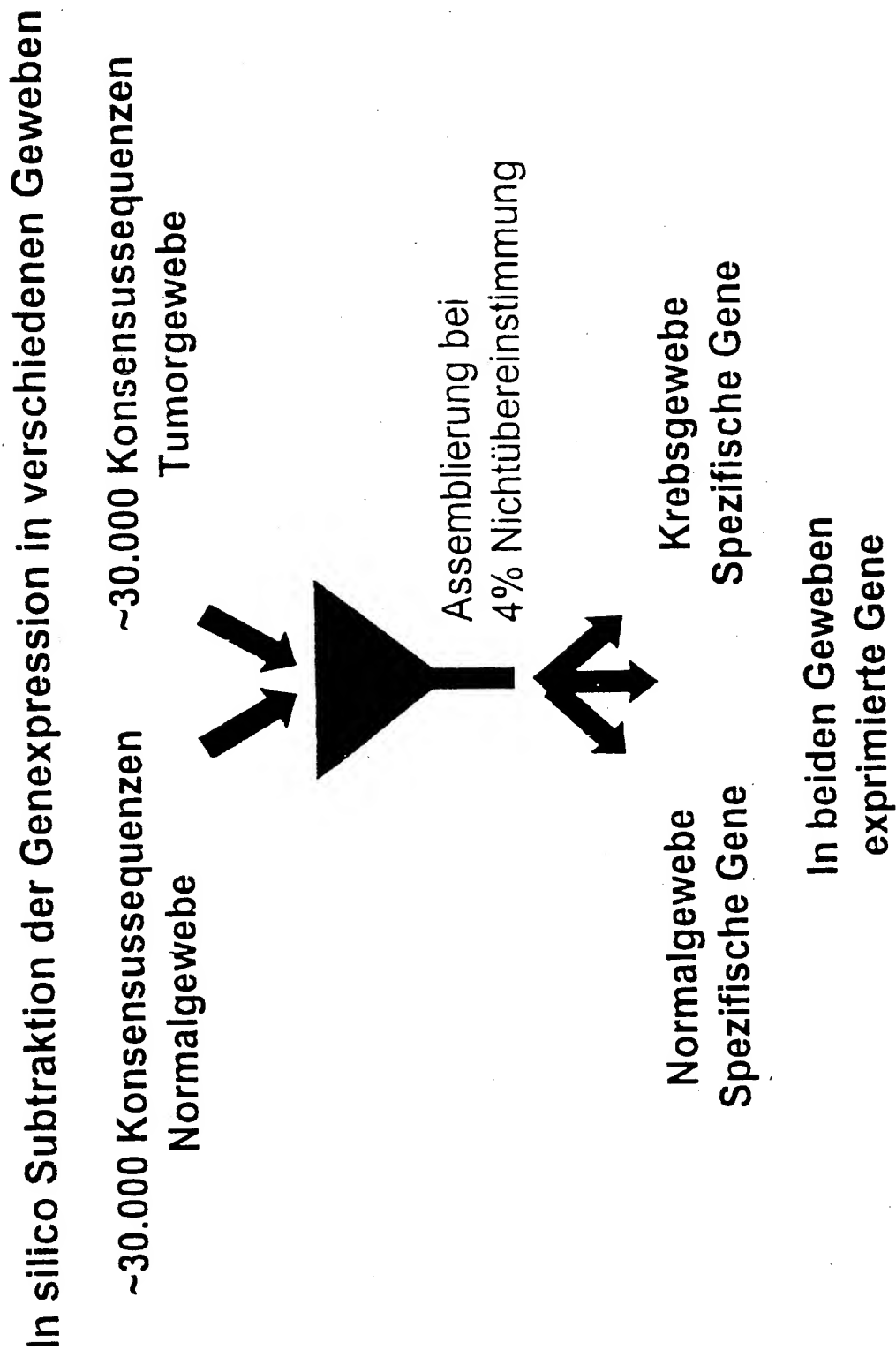


Fig. 3

8/10

Genes of interest

Determination of tissue-specific expression
via electronic Northern (INCYTE LifeSeq and
public EST databases)

Candidate genes for tumor suppressors or
tumor activators

Figure 4a

CORRECTED PAGE (RULE 91)
ISA/EP

8/10

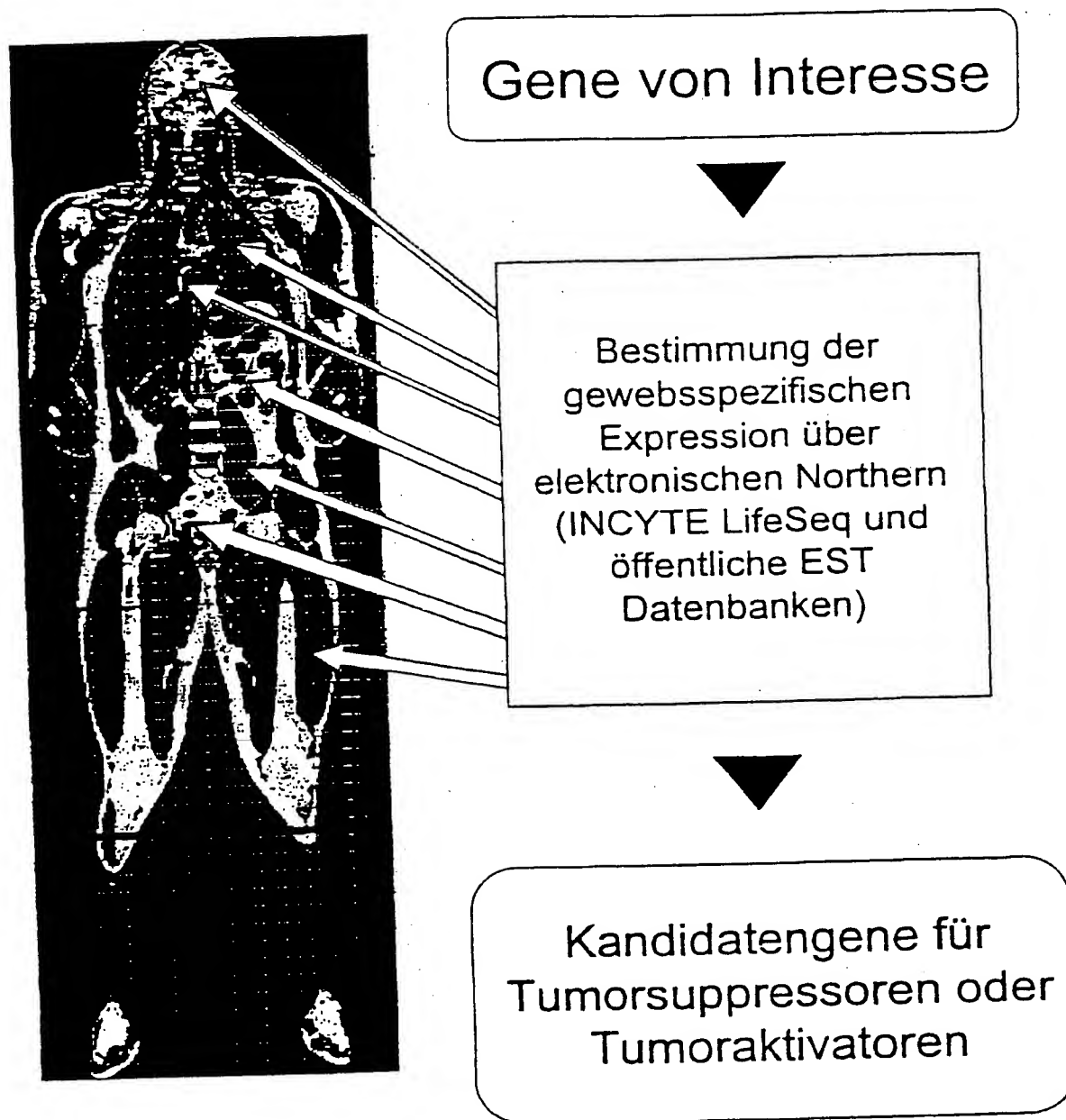


Fig. 4a

9/10

Partial cDNA sequence
e.g., EST or contig
s

...GCCTCAAGTTATC...

WHILE $C_i > C_{i-1}$

Electronic Northern Blot

Fisher's Exact Test IF H_0 EXIT

Automatic Lengthening

Consensus sequence C

...ATGTCCTAGCCTCAAGTTATCAGATGCAA...

Figure 4b

CORRECTED PAGE (RULE 91)
ISA/EP

9/10

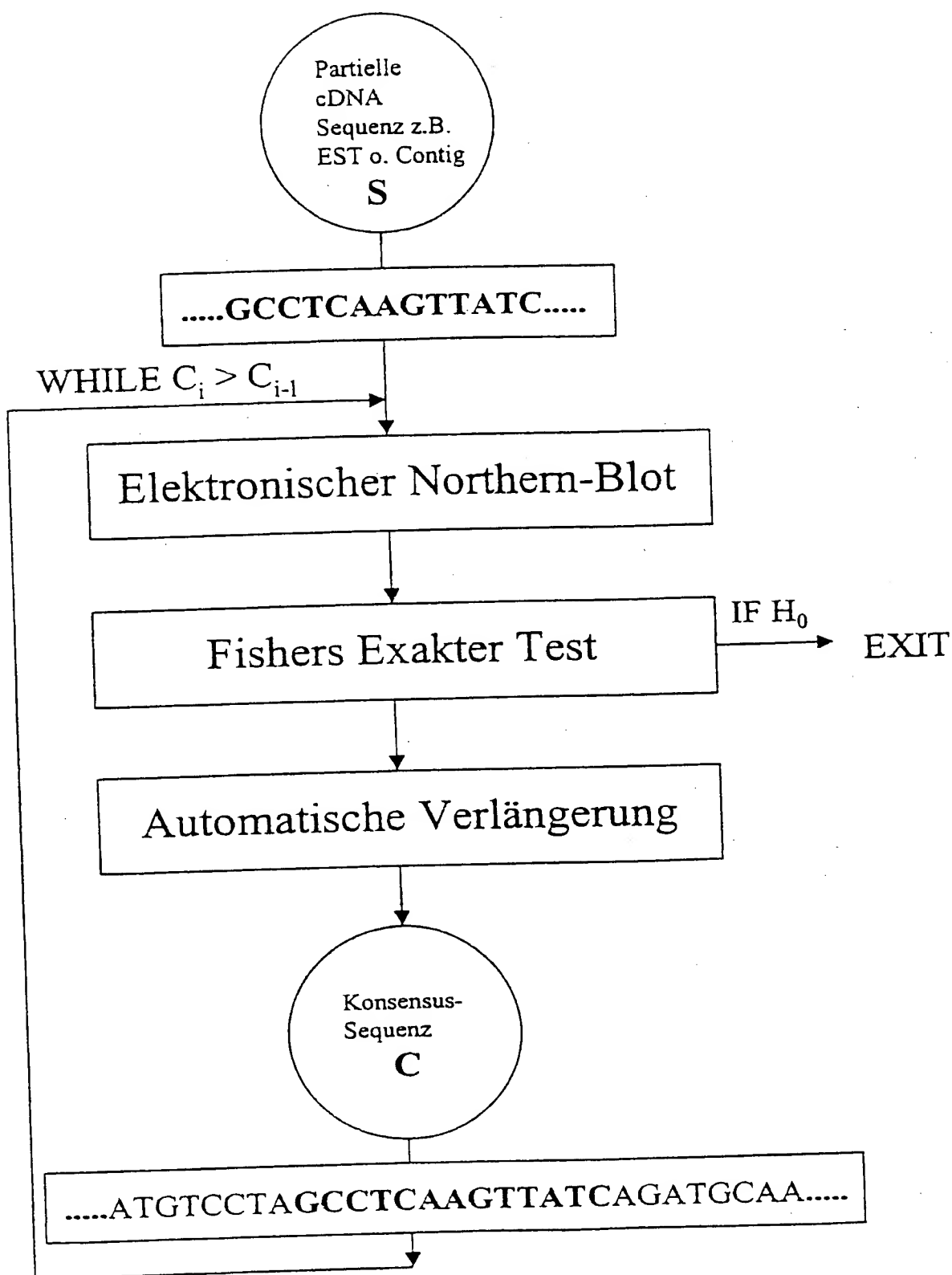


Fig. 4b

10/10

Isolation of genomic BAC and PAC clones

Chromosomal clone localization via FISH

Hybridization signal

Sequencing of clones that are located in regions that have
chromosomal deletions in prostate and breast cancer leads to
identification of candidate genes

Exon Intron

Confirmation of candidate genes by screening of
mutations and/or deletions in cancer tissues

Figure 5

CORRECTED PAGE (RULE 91)
ISA/EP

10/10

Isolieren von genomischen BAC und PAC Klonen



Chromosomale Klon-Lokalisation über FISH



Hybridisierungssignal



Sequenzierung von Klonen, die in Regionen lokalisiert sind, die chromosomale Deletionen in Prostata- und Brustkrebs aufweisen, führt zur Identifizierung von Kandidatengenen



Bestätigung der Kandidatengene durch Screening von Mutationen und/oder Deletionen in Krebsgeweben

Fig. 5